

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Deutsche Komposita zwischen Syntax und Morphologie: Ein korpusbasierter Ansatz

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/128046> since

Publisher:

Walter de Gruyter

Published version:

DOI:10.1515/9783110278439

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Linguistik – Impulse & Tendenzen

Herausgegeben von
Susanne Günthner
Klaus-Peter Konerding
Wolf-Andreas Liebert
Thorsten Roelcke

46

De Gruyter

Das Deutsche als kompositionsfreudige Sprache

Strukturelle Eigenschaften und
systembezogene Aspekte

Herausgegeben von
Livio Gaeta
Barbara Schlücker

De Gruyter

- Nishiyama, Atsuko & Jean-Pierre Koenig (2004): „What is a perfect state“. In: Benjamin Schmeiser et al. (Hgg.), *WCCFL 23 Proceedings*. Somerville, MA: Cascadia Press, 101-13.
- Portner, Paul (erscheint): „Perfect and progressive“. Erscheint in: Claudia Maienborn, Klaus von Heusinger & Paul Portner (Hgg.), *Semantics: An International Handbook of Natural Language Meaning*, Vol. 1. Berlin: Mouton de Gruyter.
- Reichenbach, Hans (1947): *Elements of Symbolic Logic*. London: Collier-McMillan.
- Rödel Michael (2007): *Doppelte Perfektbildungen und die Organisation von Tempus im Deutschen*. Tübingen: Stauffenburg.
- Rothstein, Björn (2007): *Tempus*. Heidelberg: Winter.
- Rothstein, Björn (2008): *The perfect time span. On the present perfect in German, Swedish and English*. Amsterdam: John Benjamins.
- Rothstein, Björn (2009): „Nominales Tempus im Deutschen? Zur temporalen Bedeutung von *Ex-*, *Alt-*, *Noch-* und *Jetzt-*“. *Sprachwissenschaft* 34, 435-57.
- Stechow, Arnim v. (1999): „Eine erweiterte Extended-Now Theorie für Perfekt und Futur“. *Zeitschrift für Literaturwissenschaft und Linguistik* 113, 86-118.
- Thieroff, Rolf (1992): *Das finite Verb im Deutschen. Modus – Tempus – Distanz*. Tübingen: Narr.
- Tonhauser, Judith (2006): *The Temporal Semantics of Noun Phrases: Evidence from Guaraní*. Ph.D. Thesis, Stanford University.
- Vater, Heinz (1994): *Einführung in die Zeit-Linguistik*. Hürth-Efferen: Gabel.
- Watermann, John T. (1956): „The preterite and perfect tense in German“. *Germanic review* 31, 104-14.

Deutsche Komposita zwischen Syntax und Morphologie: Ein korpusbasierter Ansatz

Livio Gaeta & Amir Zeldes

1 Einleitung

Komposita sind für die Grammatik besondere Objekte. Einerseits bilden sie morphologische Einheiten mit dem dazugehörigen Wortstatus, andererseits weisen sie, da sie aus mindestens zwei lexikalischen Basen bestehen, Strukturen auf, die i.d.R. innerhalb der Syntax ausgedrückt werden. Darüber hinaus stellt die semantische Interpretation von Komposita eine Besonderheit dar, da sie sich aus der Zusammensetzung von zwei lexikalischen Bedeutungen und einer dritten grammatischen, oft abstrakten, Relation ergibt. Im Gegensatz zu Affixen, die solche grammatischen Bedeutungen ebenfalls tragen können, tritt in der Komposition kein Morphem auf, das für die sog. ‚Wortbildungsbedeutung‘ verantwortlich gemacht werden kann. So stellt sich die Frage, woher der Bedeutungsunterschied in den Interpretationen von bspw. *Holztisch* gegenüber *Billardtisch* stammt. Welche kombinatorischen Prinzipien bestimmen die Erzeugung bzw. die Interpretation von Komposita im Allgemeinen?

In diesem Aufsatz soll aus einer synchronen Perspektive die Generierung und Interpretation deutscher Komposita empirisch untersucht werden. Wir werden sowohl Evidenz für die enge Verbindung zwischen syntaktischen Mustern und der produktiven Bildung von Komposita liefern, als auch gegen solche Auffassungen, die eine immer vorhandene syntaktische Analyse für alle Komposita annehmen. Auf der einen Seite stellen syntaktische Muster intuitiv die Grundlagen für die Interpretation der Relation zwischen Kompositionsgliedern bereit (vgl. die Paraphrasen *Tisch aus Holz* bzw. *Tisch für Billard* zu den Beispielen oben). Auf der anderen Seite bilden viele Kompositionsmuster fest verwurzelte Konstruktionen,¹ die ggf. im mentalen Lexikon gespeichert werden und als Prototypen für die produktive Erweiterung des vorkommenden Typeninventars fungieren können. Wir werden hier die These verteidigen, dass neugebildete Komposita nicht nur durch entsprechende syntaktische Strukturen

¹ In diesem Zusammenhang wird oft im Rahmen von gebrauchsbasierten Grammatikansätzen auch das englische Wort „entrenched“ verwendet, vgl. Goldberg (2006).

motiviert sind, sondern auch auf wiederkehrende morphologische Muster zurückgeführt werden sollten, die die Selektionseigenschaften der Komposita in einigen Fällen besser erklären.

Da der Entstehungsprozess neologischer Komposita nicht direkt beobachtet werden kann, werden wir dieser These anhand von Korpusdaten nachgehen. Die Untersuchung untergliedert sich dabei in zwei Bereiche. Im ersten Teil dieses Beitrags beschäftigen wir uns mit der Analyse der Wortbildungsbedeutung, die aus der Komposition resultiert. Diesbezüglich werden wir zeigen, dass man für die erfolgreiche Interpretation von Komposita auch wiederkehrende Muster in Betracht ziehen muss, die sich in großen elektronischen Korpora nachweisen lassen (vgl. Pirrelli et al. 2010). Eine solche ‚distributionelle‘ Analyse schließt jedoch nicht aus, dass gleichzeitig auch syntaktische Muster als Input für den morphologischen Kompositionsprozess dienen können. Wir werden zeigen, dass die zugrundeliegenden Strukturen größtenteils lexikalisch spezifizierte Einheiten sind und keine rein abstrakten Schemata, wie sie von Anhängern der sog. „*syntax below zero*“-Hypothese bei der morphologischen Analyse angenommen werden (vgl. Ackema 1995, Ackema & Neeleman 2004).

Anschließend gehen wir im zweiten Teil dieses Beitrags auf die speziellen Eigenschaften von Rektionskomposita ein, deren Analyse traditionell zwischen Syntax und Lexikon schwankt. Die Rolle der syntaktisch motivierten Interpretation ist in diesem Zusammenhang aufgrund der in Rektionskomposita auftretenden Argumentstrukturen von besonderer Bedeutung, da sich die Selektions- und Interpretationseigenschaften dieser Komposita oft von syntaktischen Phrasen ableiten lassen, die im „*syntax below zero*“-Ansatz unter der lexikalischen Null-Stufe als Teil der Analyse eingefügt werden.² *

Die hier vorgestellten Ergebnisse werden zeigen, dass der lexikonbasierte Ansatz zur Generierung neuer Rektionskomposita gegenüber einem rein syntaktisch motivierten Ansatz präferiert werden muss, weil er eine bessere Erklärung für deren Gebrauch liefert. Um dies zu stützen, werden wir anhand der bisher größten korpusbasierten Datengrundlage zu diesem Untersuchungsgegenstand insbesondere Fälle untersuchen, in denen Rektionskomposita gut belegt sind, deren entsprechende VPen sich kaum oder gar nicht belegen lassen.

Unsere Befunde in beiden Untersuchungsbereichen stützen ein Modell, in dem das mentale Lexikon eine zentrale Rolle bei der Erzeugung

2 In vollständig syntax-orientierten Ansätzen (bspw. bei Siebert 1999) gilt das auch für die Analyse sämtlicher anderer Komposita. Solche Analysen führen Komposition auf einen Inkorporationsprozess zurück, in dem das mentale Lexikon über die Speicherung der Basislexeme hinaus keine Rolle spielt (s. Gaeta 2010 für eine kritische Diskussion).

und der Interpretation von Komposita spielt. Das Lexikon muss in diesem Modell eine große Menge an Informationen über die Verwendung und die Frequenz von sehr vielen Typen enthalten, die ihrerseits wiederum als Vorlage für die analogische Bildung neuer Typen nach denselben Mustern fungieren. Damit wird auch die Komposition an der Schnittstelle zwischen Syntax und Morphologie Teil einer gebrauchsbasierten Grammatik, die erst seit Kurzem mithilfe von sehr großen Korpora empirisch untersucht werden kann.

2 Eine distributionelle Analyse der Wortbildungsbedeutung

Wie ergibt sich die Bedeutung eines Kompositums? Diese Frage wurde traditionell mit Bezug auf das Frege-Prinzip beantwortet, das besagt, dass sich die Bedeutung eines Kompositums kompositionell aus der Summe der Bedeutungen der einzelnen Glieder ergibt. In Anlehnung daran nimmt Becker (1992: 9) an, dass die Interpretation von Komposita im Allgemeinen aus der folgenden Generalisierung hervorgeht:

General Compounding Rule:

$\langle \langle A_N, 'A' \rangle, \langle B_N, 'B' \rangle \rangle \rightarrow \langle \langle A_N, 'B, which has some connection to A's' \rangle \rangle$

Nach dieser Ansicht ist die sog. Wortbildungsbedeutung, also die übrigbleibende Bedeutung, nachdem der Bedeutungsgehalt der einzelnen Glieder abgezogen wurde, eine generische Beziehung zwischen den Kompositionsgliedern (‘some connection’). In dieser Generalisierung bleibt jedoch die Art der semantischen Interpretation vollkommen un(ter)spezifiziert. Es wird nicht klar, woher man weiß, dass ein *Billardtisch* ein Tisch *für* das Billard ist, während ein *Holztisch* als ein Tisch *aus* Holz gedeutet wird.

Ein Versuch der genaueren Spezifizierung der möglichen semantischen Relationen, die in einem gewissen Kompositum vorkommen können, wird von Ansätzen unternommen, die semantische Schablonen, sog. *Templates*, annehmen. Diese Schablonen sollen dafür verantwortlich sein, dass eine bestimmte Wortbildungsbedeutung unter allen denkbaren bevorzugt wird (vgl. Olsen in diesem Band). Dies kann auch so verstanden werden, dass wir es mit unterschiedlichen Wortbildungsregeln zu tun haben, denn die verschiedenen Templates bilden unterschiedliche, auf der paradigmatischen Achse alternative Selektionsprozesse. Dennoch bleibt auch in solchen Ansätzen unklar, wie die korrekte Interpretation gewählt wird. Die Gesamtbedeutung bleibt auch hier un(ter)spezifiziert und muss als kontextgebunden interpretiert werden, was die ursprüngliche

Fragestellung unbeantwortet lässt, obgleich eine finite Menge an möglichen Beziehungen angenommen wird.³

Eine mögliche Lösung dieses Problems liegt in der Verfügbarkeit und Auswertbarkeit großer elektronischer Korpora. In der Literatur der letzten Jahre ist eine alte Idee aus dem klassischen amerikanischen Strukturalismus wieder hervorgetreten: Der Versuch einer Wiederbelebung der sog. „distributionellen Hypothese“ (Harris 1954). Diese beinhaltet die Annahme, dass jeder Bedeutungsunterschied in einer Sprache durch einen entsprechenden Unterschied in der Distribution, d.h. in den belegten Umgebungen einer Konstruktion, repräsentiert wird. Umgekehrt kann das distributionelle Potenzial eines Lexems ausgenutzt werden, um seine Gesamtbedeutung indirekt zu ermitteln: Elemente, die sich gleich verhalten, d.h. gleiche oder zumindest ähnliche syntaktische Kontexte präferieren, sollten auch semantisch überlappen (vgl. Miller & Charles 1991). Gleichzeitig impliziert diese Auffassung, dass die Bedeutung eines Lexems mit dem Wissen über seine Verwendungseigenschaften zusammenfällt.⁴ Trotz einiger konzeptueller sowie empirischer Schwierigkeiten hat dieser Ansatz zu interessanten Forschungsergebnissen geführt, insbesondere seitdem sehr große Textkorpora mit einer breiten Abdeckung unterschiedlicher (leider fast ausschließlich schriftlicher) Sprachregister zur Verfügung stehen (vgl. etwa Baroni & Lenci 2010).

Die Anwendung des distributionellen Ansatzes zur Ermittlung der syntaktischen und der semantischen Beziehung zwischen Kompositionsgliedern verlangt, dass der Zusammenhang zwischen eng verwandten syntaktischen Mustern erkannt wird (vgl. zum Italienischen Baroni et al. 2007, 2009). Als Arbeitshypothese würde man annehmen, dass die Interpretation des Kompositums *Holz-tisch* indirekt über die wiederkehrende Verwendung des syntaktischen Musters *Tisch aus Holz* gewonnen werden kann. Wenn hauptsächlich oder sogar ausschließlich diese syntaktische Konstellation für die Lexeme belegt wäre, könnte man behaupten, dass der Leitfaden zur Identifizierung der semantischen Verhältnisse zwischen Kopf und Modifikator in dem präpositionalen Ausdruck selbst steckt.⁵ Die wiederkehrende Verwendung dieses analytischen syntaktischen Aus-

3 Dass diese strenggenommen keine geschlossene Liste bilden, ist jedoch bekannt, vgl. Heringers (1984: 2) mehrdeutige *Fischfrau* bzw. zum Englischen den *apple-juice seat* von Downing (1977: 818).

4 Vgl. auch die bekannte Maxime des britischen Strukturalisten John Firth (1957: 11): „You shall know a word by the company it keeps“.

5 Dies bedeutet natürlich nicht, dass die Interpretation der Wortbildungsbedeutung nicht mit Weltwissen über die durch die Kompositionsglieder bezeichneten Referenten zusammenhängt, sondern lediglich, dass die wiederkehrenden syntaktischen Strukturen dieses Wissen auf der konventionalisierten sprachlichen Ebene kodieren.

drucks, die im mentalen Lexikon zusammen mit einer Repräsentation von dessen Häufigkeit gespeichert wird, würde die gewählte Interpretation des synthetischen kompositionellen Ausdrucks erklären. Bevor wir zur Besprechung der theoretischen Haltbarkeit dieser These und ihrer Konsequenzen kommen, möchten wir kurz untersuchen, inwiefern sie durch die empirischen Daten unterstützt wird.

Da nur das sehr häufige Vorkommen einer bestimmten syntaktischen Struktur dafür verantwortlich gemacht werden kann, dass diese auch zur Interpretation von entsprechenden, ggf. viel seltener vorkommenden, im mentalen Lexikon fest verankerten Komposita beiträgt, benötigen wir ein möglichst großes Korpus, um diesen Zusammenhang überprüfen zu können. Zu diesem Zweck verwenden wir hier einen Ausschnitt aus dem deWaC-Korpus (deutsch: *Web as Corpus*), das Daten von deutschsprachigen Web-Seiten enthält. Dieses Subkorpus, deWaC01, enthält insgesamt ca. 170 Millionen Token, etwa ein Zehntel des gesamten deWaC-Korpus: Insgesamt ca. 1,7 Milliarden mit Wortarten und Lemmata automatisch annotierten Token (vgl. Baroni et al. 2009). Obwohl dieses Korpus hinsichtlich der enthaltenen Textsorten nicht ausgewogen ist, ist es aufgrund seiner Größe eine unvergleichlich wertvolle Ressource für die Sammlung ausreichender Belege bei verhältnismäßig selten vorkommenden Phänomenen. Der ausbleibende Beleg einer syntaktischen Struktur für dieses Korpus bedeutet natürlich nicht, dass diese grundsätzlich nicht existiert. Man würde aber erwarten, dass syntaktische Strukturen, die als Muster für die Interpretation von Komposita verwendet werden können, einigermaßen gut belegt sein sollten.⁶ Im Rahmen der Untersuchung wurden automatisch alle Wörter mit mindestens 15 Buchstaben als mögliche Komposita heuristisch aus dem Korpus extrahiert, zusammen mit ihren jeweiligen Frequenzen. Die Frequenzliste wurde anschließend in vier Frequenzgruppen eingeteilt: sehr seltene Typen, sog. Hapax Legomena (Frequenz = 1), seltene Typen (Frequenz = 2-10), häufige Typen (Frequenz = 11-163) und sehr häufige Typen (> 163). Anschließend wurden 500 Komposita jeder Frequenzgruppe manuell kontrolliert, um automatisch extrahierte Nicht-Komposita auszuschließen. Nach der Beseitigung von lediglich 109 falschen Treffern, einer Fehlerquote von ca. 5 %, blie-

6 Um die Größe des Korpus vor dem Hintergrund bisheriger Arbeiten zu relativieren, sei erwähnt, dass Baayen & Renouf (1996) bei ihrer Untersuchung einiger englischen Affixe aus 40 Monaten der täglichen Ausgabe der Tageszeitung *Times* (1989-1993) ein Korpus von etwa 80 Millionen Token zusammenstellen konnten. Das Teilkorpus deWaC01 alleine ist mehr als zweimal so groß wie das *Times*-Korpus, und das gesamte Korpus, das in Abschnitt 3 verwendet wird, entsprechend über zwanzigmal so groß.

ben 1891 Komposita für die folgende Untersuchung.⁷ Die Einteilung in verschiedene Frequenzklassen hat zum Ziel, den Parameter Frequenz zu kontrollieren, da wir davon ausgehen, dass sich unterschiedlich frequente Typen bei der Erzeugung und Interpretation ihrer Bedeutung auch unterschiedlich verhalten.

Um die Relevanz der Korpusdaten für die Frage nach den Erzeugungstypen neuer Komposita einschätzen zu können, wurde eine Kontrollstudie durchgeführt, in der wir die im Korpus belegten Komposita jeder Frequenzgruppe mit den Vorkommen in einem Standardwörterbuch abgeglichen haben (Daten aus der Online-Version des Duden, www.duden.de). Die folgende Tabelle fasst die Ergebnisse zusammen:

Frequenzklassen		Komposita	
I	Hapax Legomena	7	0,4 %
II	2-10	4	0,2 %
III	11-163	36	1,9 %
IV	> 163	199	10,5 %
	Tot.	246	13,0 %

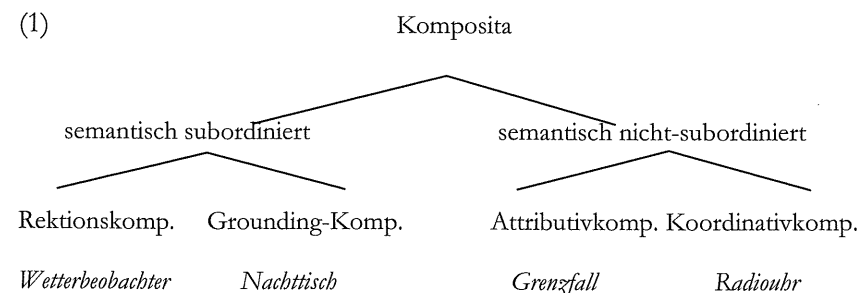
Tab. 1: Anzahl der im Duden vorkommenden Komposita

Wie aus Tabelle 1 hervorgeht, ist nur ca. ein Zehntel der untersuchten Komposita im Duden belegt und erwartungsgemäß machen die Komposita der vierten Klasse dabei den Löwenanteil aus. Bei den im Duden nicht vorkommenden Komposita aus unserem Datensatz vermuten wir, dass deren Bedeutung kompositionell erschließbar ist, sodass sie lexikographisch nicht erfasst werden mussten.

Um die entsprechenden syntaktischen Muster genauer untersuchen zu können, wurden die Komposita als nächstes nach semantischen Typen kategorisiert. Zu diesem Zweck wurde die von Baroni, Guevara & Pirrelli (2007, 2009) entwickelte Typologie der semantischen Verhältnisse zwischen Kopf und Modifikator innerhalb eines Kompositums verwendet, die auch mit anderen Klassifizierungsversuchen vergleichbar ist (vgl. u.a. Bisetto & Scalise 2005 für einen Überblick). Nach ihrer Taxonomie werden die vier Grundtypen in (1) angenommen:⁸

⁷ Die Extraktion der Komposita wird in Gallo (2006-2007: 61-65) im Einzelnen beschrieben. Wir möchten Frau Gallo für ihre Arbeit bei der Extraktion und der empirischen Untersuchung der deWaC-Daten an dieser Stelle herzlich danken.

⁸ Wir lassen dabei die im Deutschen seltene Untergruppe der Possessivkomposita außer Acht (s. Baroni et al. 2007).



Wir werden die einzelnen Eigenschaften dieser Grundtypen hier nicht besprechen. Es sei nur darauf hingewiesen, dass der Unterschied zwischen den Rektionskomposita und den sog. Grounding-Komposita⁹ darin besteht, dass der Nichtkopf bei den Rektionskomposita eine Argumentstelle des Kopfexems sättigt. Der Kopf ist in diesen Fällen daher meistens deverbalen Natur. Darüber hinaus ist die Unterscheidung zwischen attributiven und koordinativen Komposita nicht immer leicht zu treffen, je nachdem, ob der Nicht-Kopf als Attribut des Kopfes oder als Bestandteil der Extension des Kompositums gedeutet wird. So könnte bspw. ein Kompositum wie *Affenmensch* theoretisch sowohl attributiv als auch koordinativ verstanden werden (vgl. Olsen 2001, Gaeta & Ricca 2009): Es könnte sich um einen Menschen handeln, der auch ein Affe ist, oder um einen Menschen, der einem Affen ähnlich ist bzw. Affenmerkmale aufweist (auch wenn letztere Lesart sicherlich plausibler erscheint). Die eindeutigen Fälle von Koordinativkomposita sind in erster Linie die rein additiven wie *Nordrhein-Westphalen*, *Österreich-Ungarn* usw., die in der altindischen Klassifizierung auch *Dvandva* genannt werden (vgl. Bauer 2001).

Allerdings spielt diese Unterscheidung eine untergeordnete Rolle für unsere Stichprobe von 1891 Komposita, da beide Typen zusammen nur ca. 3 Prozent der Belege darstellen (44 bzw. 2 % attributive und 20 bzw. 1 % koordinative Fälle).¹⁰ Erwartungsgemäß bilden Grounding-

⁹ Auch Root Compounds genannt, vgl. Roeper & Siegel (1978). Wir ziehen den von Baroni, Guevara & Pirrelli (2007) verwendeten Begriff Grounding vor, weil er besser als der englische Begriff „Root“ auf die Haupteigenschaft dieser Komposita hinweist, dass der Modifikator den Kopf kontextualisiert bzw. näher bestimmt (d.h. *the modifier „grounds“ the head*). Für die hier verwendete Terminologie s. Baroni et al. (2007).

¹⁰ Dass die große Mehrheit der Komposita eine Subordination ausdrückt, entspricht zwar der Erwartung. Man beachte aber, dass dies nicht unbedingt der Fall sein muss. In einer Sprache, in der Komposition viel weniger häufig verwendet wird, wie dem Italienischen, sieht die Verteilung deutlich anders aus. Wie Baroni et al. (2009) zeigen, bilden in einer Stichprobe von 252 aus einem ähnlichen Web-Korpus gezogenen italienischen Komposita die nicht-subordinierten Fälle etwa 30 % der Typen, während Rektionskomposita etwa 20 %

Komposita den Löwenanteil (1421 bzw. 76 % der Typen), während die Rektionskomposita viel weniger häufig sind (406 bzw. 21 % der Typen). Der Unterschied zwischen diesen beiden Typen von subordinierten Komposita kann teilweise darauf zurückgeführt werden, dass die deverbalen Köpfe, die an der Bildung der meisten Rektionskomposita beteiligt sind, viel seltener vorkommen als die nicht-deverbalen Köpfe.

Auf Grundlage dieser Daten wurde die vergleichende distributionelle Analyse nach einem ursprünglich von Lauer (1995) vorgeschlagenen Verfahren angewendet (vgl. Pirrelli et al. 2010 für eine kritische Diskussion). Dieses Verfahren nimmt als Ausgangspunkt die semantische Entsprechung zwischen den Wortbildungsbedeutungen innerhalb der Komposita und den syntaktischen Konstruktionen an, die solche Bedeutungen auf der Satzebene kodieren, wie z.B. Präpositionalphrasen. Das Korpus wird nach sog. Konnektorenmustern (*Connector Patterns*) durchsucht, die sich aus der Entsprechung zwischen dem jeweiligen Kompositum und seiner syntaktischen Paraphrase ergeben, wie bspw. *Computerraum* ↔ *Raum für Computer*. Da *Computerraum* ein Grounding-Kompositum ist, erwartet man nach diesem Ansatz eine statistisch signifikante Korrespondenz zwischen dem Konnektorenmuster „X *für* Y“ und der Wortbildungsbedeutung „Grounding“ in einigen Komposita der Form Y-X. Diese Hypothese können wir nun mithilfe unseres Korpus überprüfen.¹¹ Die Untersuchung basiert auf der Untermenge der nur aus N-N-Paaren bestehenden, insgesamt 1723 Komposita mit drei Typen von Konnektorenmustern, die funktional sowie semantisch deutlich voneinander zu unterscheiden sind:

- (2) a. Koordinierende Konjunktionen: *und, oder, bzw., als auch, wie, als, entweder ... oder, weder ... noch*;
 b. Definiter Artikel im Genitiv bzw. Dativ;
 c. Präpositionen: *für, aus, gegen, bei, mit, von, zu, an, auf, in, über, vor, zwischen, unter, nach, um, ohne, durch*.

der Typen und die Grounding-Komposita „nur“ ca. 47 % der Typen darstellen. Der Unterschied in der Anzahl der Grounding-Komposita kann als Symptom für den unterschiedlichen Status betrachtet werden, der der Komposition in den zwei Sprachen zukommt: Im Italienischen scheint der rein subordinierte Typ in Abwesenheit einer argumentartigen Relation viel weniger häufig auftreten zu können. Zu diesem typologischen Thema besteht noch erheblicher Forschungsbedarf auf Basis möglichst großer vergleichbarer Korpora.

- 11 Dabei sei darauf hingewiesen, dass wir aufgrund der Ergebnisse von Lauer (1995) eine erhebliche Anzahl von Belegen für wiederkehrende syntaktische Muster erwarten, da das hier verwendete Korpus mit ca. 170 Millionen Token weitaus größer ist, als das Korpus, das Lauer für die Untersuchung englischer Komposita ursprünglich zur Verfügung stand, mit nur ca. 8 Millionen Token.

Für jedes Paar wurde dann die statistische Signifikanz der Vorkommenshäufigkeit der Lexeme mit den jeweiligen Konnektoren-mustern mithilfe der *Log-Likelihood Ratio* (vgl. Baroni et al. 2009) berechnet.¹²

In Tabelle 2 ist für die einzelnen Konnektorenmuster und die einzelnen Komposita-Klassen (Attributive Komposita = A; Koordinative Komposita = K; Rektionskomposita = R; Grounding-Komposita = G) das Verhältnis der Kopf-Modifikator-Paare dargestellt, die eine signifikante Zusammentreffenshäufigkeit ($p < 0.01$) in Verbindung mit dem gegebenen Konnektorenmuster aufweisen:

	A	K	R	G
(2a)	50 %	59 %	11 %	19 %
(2b)	7 %	7 %	35 %	18 %
(2c)	43 %	34 %	54 %	63 %

Tab. 2: Verhältnis der signifikant vorkommenden Kopf-Modifikator-Paare in Verbindung mit den jeweiligen Konnektorenmustern

Wie aus der Tabelle ersichtlich ist, korrelieren die Konnektorenmuster in (2a) mit den attributiven und den koordinativen Komposita: In ca. 50 % bzw. 60 % der Fälle stimmen die durch die dargestellten Konnektorenmuster verbundenen Wortpaare mit den Wortpaaren überein, die auch bei den attributiven bzw. koordinativen Komposita vorkommen. Diese Übereinstimmung ist bei den Rektions- bzw. Grounding-Komposita hingegen mit jeweils 11 % bzw. 19 % viel geringer. Das bedeutet, dass die in (2a) angegebenen Konnektoren viel weniger häufig die Kopf-Modifikator-Paare verbinden, die bei subordinierten Komposita vorkommen. Erwartungsgemäß werden die Kopf-Modifikator-Paare der subordinierten

- 12 Dabei wurde die beobachtete und erwartete Häufigkeit der Beziehung Kopf-Modifikator-Konnektor auf folgende Weise berechnet (p = Wahrscheinlichkeit; H = Kopf (Head); M = Modifikator; CONN = Konnektor):

Beobachtete Häufigkeit: $\text{count}(H \text{ CONN } M)$

Erwartete Häufigkeit: $p_{\text{ind}}(HM) \times N_{\text{CONN}}$

Wobei: $p_{\text{ind}}(HM) = p(H) \times p(M)$

$$p(H) = \frac{\text{count}(H)}{N_{\text{CORPUS}}}$$

$$p(M) = \frac{\text{count}(M)}{N_{\text{CORPUS}}}$$

$$N_{\text{CONN}} = \text{count}(\text{NOUN CONN NOUN})$$

Komposita aber viel häufiger durch typische subordinierende Konnektorenmuster verbunden, wie die Präpositionen.

Darüber hinaus werden die bei koordinativen Komposita vorkommenden Kopf-Modifikator-Paare deutlich seltener durch Präpositionen verbunden. Dabei sollte man beachten, dass bei den attributiven Komposita der Anteil der präpositionalen Konnektorenmuster im Vergleich zu den koordinierenden Konnektoren relativ hoch bleibt, etwa 43 %. Die empirischen Ergebnisse sind hier daher nicht so eindeutig wie in den anderen Fällen. Das könnte jedoch damit zusammenhängen, dass bei solchen attributiven Komposita Modifikatoren wie *Grund-* bzw. *Hilf-* häufiger auftreten, die auch in komplexen Präpositionen erscheinen, wie *auf Grund (von)*, *mit Hilfe (von)* usw., was dazu führt, dass falsche Entsprechungen von Kopf+Präp.+Modif. automatisch erkannt wurden.¹³

Auch der Unterschied zwischen subordinierten und nicht-subordinierten Komposita bei dem zweiten Konnektorenmuster (2b) ist auffällig: Jeweils 7 % gegenüber 35 % bzw. 18 %, obwohl die niedrigeren Werte in diesem Fall vermutlich auf die reduzierte Verwendung des reinen Artikels als syntaktisches Bindemittel gegenüber den Präpositionen zurückzuführen sind. Jedenfalls scheint dieses Konnektorenmuster ein guter Indikator für die Unterscheidung von Rektions- und Grounding-Komposita zu sein, wobei die Grounding-Komposita eher die anderen zwei Konnektoren und insbesondere die Präpositionen bevorzugen.

Betrachtet man die Verteilung der einzelnen Präpositionen in Tabelle 3, die mit den Kopf-Modifikator-Paaren erscheinen, so wird der Unterschied zwischen Rektions- und Grounding-Komposita noch deutlicher:

13 Ein ähnliches Problem wurde auch im Italienischen beobachtet, wo Ausdrücke, die eine Präposition umfassen, wie *in base* ‚auf Grund‘ oder *al limite* ‚grenzwertig‘ eine fälschlicherweise hohe Proportion von Präpositionen bei den attributiven Kompositionsgliedern verursachen dürften (vgl. Baroni et al. 2009).

Präposition	R	G
<i>An</i>	5%	7%
<i>Auf</i>	5%	5%
<i>aus</i>	4%	5%
<i>bei</i>	4%	7%
<i>durch</i>	2%	4%
<i>für</i>	14%	19%
<i>gegen</i>	1%	1%
<i>in</i>	10%	16%
<i>mit</i>	7%	7%
<i>nach</i>	4%	2%
<i>ohne</i>	1%	-
<i>über</i>	4%	5%
<i>um</i>	-	1%
<i>unter</i>	-	-
<i>von</i>	29%	14%
<i>vor</i>	2%	-
<i>zu</i>	8%	9%
<i>zwischen</i>	1%	-

Tab. 3: Anteil der signifikant vorkommenden Kopf-Modifikator-Paare bei Rektions- und Grounding-Komposita für die einzelnen Präpositionen

Die in Rektionskomposita vorkommenden Kopf-Modifikator-Paare weisen eine eindeutige Präferenz zur Verbindung durch *von* auf, im Gegensatz zu den Grounding-Komposita, bei denen einige Präpositionen häufig erscheinen, aber insbesondere *für* und *in*, die typischerweise adjunktive Bedeutungen ausdrücken. Da die Präposition *von* zum Ausdruck von syntaktischen Verhältnissen innerhalb der Nominalphrase unmarkiert ist und als Alternative zum Kasusmarkierten Artikel oft auftritt, bestätigt dieser Befund unsere Erwartung, dass die Grounding-Komposita ein semantisch differenzierteres Bild aufweisen (wie auch Beckers ‚some connection‘ vermuten lässt). Die unterschiedlichen Wortbildungsbedeutungen, die durch die nicht-deverbale Komposition ausgedrückt werden, lassen sich mit der generischen Präposition *von* weniger gut realisieren, wäh-

rend die spezifischeren Präpositionen *für* und *in* die semantische Relation expliziter machen.

Die oben dargestellten Ergebnisse scheinen zunächst die Ausgangshypothese zu unterstützen. Die Interpretation der Komposita, d.h. deren ‚Wortbildungsbedeutung‘, zeigt eine erhebliche Korrelation mit den wiederkehrenden syntaktischen Mustern, in denen die entsprechenden Kopf-Modifikator-Paare vorkommen. Einerseits wird die Unterscheidung zwischen Rektions- und Grounding-Komposita innerhalb der traditionellen Determinativkomposita durch diese distributionelle Analyse bestätigt. Andererseits wird durch die distributionelle Analyse das jeweilige semantische Profil der unterschiedlichen Kopf-Modifikator-Paare mithilfe ihres syntaktischen Verhaltens identifiziert, wenngleich klar bleibt, dass die Zuordnung der semantischen Klassen zu den belegten Konnektoren nicht eindeutig ist und lediglich Tendenzen offenlegen kann.

Aus diesen Ergebnissen kann man jedoch noch nicht schließen, dass eine rein syntaktische Analyse im Sinne der transformationellen Grammatik für die Bildung der Komposita bevorzugt werden sollte.¹⁴ Die belegten Kopf-Modifikator-Paare können aus semantischen Gründen neben den Komposita existieren, ohne dass Letztere direkt von den syntaktischen Mustern abgeleitet werden.¹⁵ Obwohl die durchgeführte distributionelle Analyse eine empirische Beziehung zwischen Komposita und syntaktischen Phrasen zeigt, findet sich auch empirische Gegenevidenz in der näheren Betrachtung der belegten Lexeme, die in der Komposition bzw. in syntaktischen Mustern erscheinen. Der nächste Abschnitt widmet sich v.a. den speziellen Eigenschaften der Rektionskomposita, die den engsten Bezug zu syntaktischen Mustern erwarten lassen.

14 Ein Beispiel für diese Auffassung findet sich in Siebert (1999), der davon ausgeht, dass jedes deutsche Kompositum durch Kopf-zu-Kopf-Bewegung und Inkorporation der regierten DP zustande kommt. Dementsprechend erhält ein Kompositum wie *Härtegrad* die strukturelle Analyse in (i):

(i) [*Härtegrad* [_i [_{t_i}]_{NP}]_{DP}]_{NP}

(ii) *der Härtegrad* _{t_i} *des Wassers*

Die Analyse soll die Tatsache erklären, dass in einer DP wie (ii) ein Rektionsverhältnis zwischen *Härte* und *Wasser* besteht und nicht etwa zwischen dem Kompositionskopf *Grad* und dem Genitivattribut *des Wassers* (vgl. Gaeta 2010 für eine kritische Diskussion).

15 Bei einer vermeintlich de-syntaktischen Generierung der Komposita müssen wir auch noch die phylogenetische von der ontogenetischen Perspektive trennen: Es kann sein, dass Komposition in den natürlichen Sprachen diachron auf die Grammatikalisierung von syntaktischen Mustern zurückgeführt werden muss (vgl. Gaeta 2008). Ob der synchrone Generierungsprozess, der für die Produktion bzw. Analyse der Komposita in einer konkreten Sprechsituation verantwortlich ist, eine Ableitung von einer syntaktischen Phrase beinhaltet, ist eine davon unabhängige Frage.

3 Lexikalische Muster bei Rektionskomposita

Die Frage nach der Rolle der Syntax innerhalb eines Grammatikmodells kann hier nicht im Detail besprochen werden. Rektionskomposita stellen jedoch einen sehr wichtigen Untersuchungsgegenstand für die Überprüfung des Zusammenspiels zwischen Syntax und Morphologie dar. Ihr problematischer Status wird besonders deutlich im Falle der sog. Zusammenbildungen wie etwa *Totsagung*, *Verhaftnehmung* usw., in denen der Kopf **Sagung*, **Nehmung* außerhalb der Komposition für gewöhnlich nicht vorkommt. In diesen Fällen lässt sich das Kompositum nicht durch die Zusammensetzung zweier lexikalischen Einheiten herleiten, sondern muss auf eine Verbalphrase wie [*tot sagen*]_{VP} bzw. [(*in*) *Verhaft nehmen*]_{VP} zurückgeführt werden.¹⁶ Weitere Probleme entstehen durch das Vorkommen von sog. Klammerparadoxen wie z.B. *saure Gurkenzeit*, wo die strukturelle Analyse (in eckigen Klammern) der semantischen Interpretation (in Spitzeklammern) widerspricht: [_{<saure [Gurken>zeit]}].

Solche Beispiele sprechen scheinbar dafür, dass Einheiten, die auf der syntaktischen Ebene erzeugt werden, als Input für Wortbildungsprozesse dienen, die normalerweise im Lexikon angesiedelt werden. Solche Phänomene können auch dadurch erklärt werden, dass im Lexikon bzw. in der Wortbildung ähnliche generative Prinzipien wie in der Syntax gelten, was sich mit dem o.g. Begriff „*syntax below zero*“ zusammenfassen lässt (Ackema 1995, Ackema & Neeleman 2004). Das Vorkommen von Argumentstrukturen bei deverbalen Rektionskomposita, die der verbalen Argumentstruktur stark ähneln, legt einen solchen Erklärungsansatz nahe.

Um das Erklärungspotenzial der syntaktischen These zu überprüfen, haben wir aus dem gesamten deWaC-Korpus Rektionskomposita extrahiert, die als Kopf ein Agensnomen auf *-er* haben, wie etwa *Taxifahrer*, *Wetterbeobachter* usw., um diese dann mit den entsprechenden Objekt-Verb-Paaren (OV) zu vergleichen.¹⁷ Mit den Korpus-Daten können wir nun die Hypothese überprüfen, dass alle Rektionskomposita durch OV-Muster

16 Es sei denn, die zwei möglichen aber unbelegten Ableitungen °*Sagung* bzw. °*Nehmung* werden als Zwischenschritte angenommen, was zu anderen Schwierigkeiten führen würde.

17 Das automatisch lemmatisierte und POS-annotierte Korpus wurde nach folgendem syntaktischen Muster durchsucht: KONJ SUBJ (REFL) DP_{NOM} (ADJUNKTE) DP_{AKK} VVFIN. Verb- und Objektlexeme wurden anschließend mit Kompositionsgliedern anhand von Substring-Matching und einfachen Umlaut- bzw. Schwatigungsregeln sowie einer maschinellen Entfernung von Fugenelementen gekoppelt. Die Beschränkung auf Verbletztsätze mit einer subordinierenden Konjunktion ermöglicht die sonst schwierige automatische Erkennung des Akkusativobjekts. Die Treffermenge ist dennoch so groß, dass wir von einer sehr guten Abdeckung des Objektinventars eines jeden Verbs ausgehen (vgl. Gaeta & Zeldes Ms. für eine detailliertere Darstellung und Diskussion dieser Untersuchung).

motiviert sind. Eine fehlende syntaktische Belegung für Lexeme, die häufig als Kompositionsglieder vorkommen, würde einen Hinweis dafür liefern, dass ein anderer Mechanismus im Spiel sein kann, v.a. wenn solche Fälle systematische Reihen bilden sollten. Einige Abweichungen von der erwarteten Korrespondenz wären jedoch nicht verwunderlich, wie bspw. rein lexikalisierte, nicht (mehr) kompositionelle Komposita wie *Schriftsteller* (synchron vermutlich nicht < *Schrift* + *stellen*), oder bei Komposita, die zwar auf den ersten Blick wie Rektionskomposita aussehen, bei denen tatsächlich aber kein echtes argumentartiges Verhältnis zwischen dem deverbale Kopf und dem Modifikator vorliegt, wie bei *Sonntagsfahrer* oder *Radarbeobachter*.

Wir gehen zunächst der Frage nach, ob es eine starke Korrespondenz zwischen Rektionskomposita und OV-Paaren gibt. Die gefundene Menge der Verbalobjekte umfasst ca. 925.000 verschiedene Verb-Objekt-Kombinationen, von denen ca. 40.000 häufiger als 5-mal vorkommen und 3000 unterschiedliche Verbtypen aufweisen. Wir finden insgesamt 5266 gut belegte Komposita (Frequenz > 5), deren Kopflexeme als transitive Verben belegt sind. Von diesen gut belegten Fällen finden sich 4538 Lexempaare (86 %) auch in der Liste der Verbalobjekte. Diese Ergebnisse unterstützen die oben aufgestellte These nur teilweise: Zwar wird die absolute Mehrheit der Rektionskomposita durch ein OV-Muster widerspiegelt, andererseits sind Rektionskomposita verglichen mit den viel häufiger belegten OV-Paaren ein seltener vorkommendes Phänomen. Wir finden über 700 gut belegte Rektionskomposita ohne syntaktische Entsprechung, trotz der enormen Größe des Korpus und der Seltenheit dieses Phänomens. Diese müssen daher genauer betrachtet werden: Welche Köpfe sind für diese Komposita charakteristisch? Bilden sie eine ungeordnete zufällige Menge oder lassen sich Muster erkennen? Welche Köpfe bilden überhaupt viele Rektionskomposita? Sind es dieselben Lexeme, die als Verben viele Objekte haben? Und korreliert die Frequenz der OV-Paare mit der der entsprechenden Rektionskomposita?

Wir beantworten zunächst die letzten beiden Fragen: Statistisch gesehen besteht keine signifikante Korrelation zwischen der Frequenz der Rektionskomposita und der Frequenz der korrespondierenden OV-Paare ($r^2=0.0007$). Außerdem haben viele Modifikatoren bei der Komposition nicht unbedingt viele Verbalobjekte in VPen für ein Kopflexem zur Folge (eine überraschend schwache Korrelation von $r^2=0.2588$). Das bedeutet, dass es Köpfe gibt, die tendenziell viele Verbalobjekte haben, demgegenüber aber nur wenige Rektionskomposita bilden, und umgekehrt.

Zwar widerspricht dieser Befund nicht direkt der These, dass Rektionskomposita durch syntaktische VP-Muster motiviert sind, da diese auf

einer kompetenzorientierten Auffassung von Grammatik beruht, die performanzbasierte Gegebenheiten wie Frequenz nicht berücksichtigt. Ein kompetenzorientierter Ansatz kann die o.g. Beobachtungen jedoch nicht erklären und bietet keine Möglichkeit an, die syntaktische Generierung der Rektionskomposita empirisch zu untersuchen.

Noch problematischer für die syntaktische These sind die ca. 700 häufig ($f > 5$) belegten Fälle von Rektionskomposita, die keinem VP-Muster entsprechen, was bei der Größe des vorliegenden Korpus bedeutsam erscheint. Teile von dieser Untermenge fehlen sicherlich nur zufällig, entweder weil sie in anderen, hier nicht untersuchten Satztypen erscheinen oder weil die Lexeme so selten sind, dass selbst dieses große Korpus sie nicht enthält. Eine qualitative Analyse dieser Fälle lässt jedoch vermuten, dass es einen großen Anteil an häufigen Zusammensetzungen gibt, die keine syntaktische Realisierung aufweisen.

Rektionskomposita	OV	f(RK)
<i>Krankheitserreger</i>	? <i>Krankheit erregen</i>	5481
<i>Wirtschaftsprüfer</i>	? <i>Wirtschaft prüfen</i>	5347
<i>Mobilfunkbetreiber</i>	? <i>Mobilfunk betreiben</i>	3207
<i>Handelsvertreter</i>	? <i>Handel vertreten</i>	3009
<i>Automobilhersteller</i>	<i>Automobile herstellen</i>	2923
<i>Reiseleiter</i>	? <i>Reise leiten</i>	2584
<i>Medienvertreter</i>	<i>Medien verrete</i>	2506
<i>Naturforscher</i>	? <i>Natur forschen</i>	2430
<i>Konkursverwalter</i>	? <i>Konkurs verwalten</i>	2146
...		
<i>Staubsaugervertreter</i>	<i>Staubsauger vertreten</i>	116

Tab. 4: Als OV-Paar nicht belegte häufige Rektionskomposita

In Tabelle 4 werden einige häufige Rektionskomposita angegeben (s. die Frequenz in der letzten Spalte), die in unserem Korpus auffälligerweise keinem OV-Paar entsprechen. Diese stellen lexikalisierte Ausdrücke dar, die die Verwendung des entsprechenden Syntagmas blockieren oder zumindest sehr markiert erscheinen lassen: Man spricht bspw. von *Krankheitserregern* und *Konkursverwaltern*, vermeidet jedoch offenbar eine Verbalisierung dieser Kombinationen. Andere Fälle hängen mit veralteten Lesarten der Kompositionsglieder bei Erhalt des ganzen Kompositums

zusammen wie im Fall des sehr häufigen Kompositums *Arbeitnehmer* (über 135.000 Belege), das im Korpus neben anderen Komposita wie *Forschungsnehmer*, *Projektnehmer*, *Vertragsnehmer* usw. vorkommt. Die entsprechenden OV-Paare erscheinen kaum mit der relevanten Bedeutung: ?*Arbeit* / ?*Forschung* / ?*Projekt* / ?*Vertrag nehmen*, obwohl sie in der Barockzeit ggf. belegt sind: Vgl. GWB s.v. *nehmen*: *wann einer [...] solche arbeit auff sich nehmen wolt* (Schuppius 10). Oder das Kompositum erhält alte Formen, die außerhalb des Kompositums nicht mehr verwendet werden (z.B. *Automobilhersteller* aber synchron üblicherweise *Auto*, nicht *Automobil*). In beiden Fällen muss man davon ausgehen, dass diese Einheiten im mentalen Lexikon gespeichert und nicht synchron von syntaktischen Mustern abgeleitet werden. Angesichts der Häufigkeit dieser Komposita würde man in einem gebrauchsbasierten Ansatz vielmehr davon ausgehen, dass der Erscheinung eines Syntagmas wie bspw. *X erregt eine Krankheit* gerade die nominale Form zugrunde liegt.

Als letztes stellt sich die Frage, ob es auch nicht lexikalisierte Rektionskomposita gibt, die von syntaktischen Phrasen schlecht abgeleitet werden können. Dazu müssen wir einen Blick auf die selteneren Rektionskomposita werfen, die in der Darstellung oben außer Acht gelassen wurden: Die Komposita mit einer Frequenz < 5, oder noch besser, die nur einmal belegten sog. Hapax Legomena. Letztere werden in vielen Ansätzen untersucht, um das Verbreitungsbild einer Konstruktion bei produktivem Gebrauch zu untersuchen (vgl. u.a. Baayen 2001, Gaeta & Ricca 2006, Zeldes 2009). Obwohl es an sich nicht weiter verwunderlich ist, dass es seltene Lexempaare gibt, die per Zufall nur als Rektionskomposita belegt sind, stellt man hier fest, dass es einige Kopflexeme gibt, die für einen erheblichen Anteil der Hapax Legomena verantwortlich sind. Die Köpfe mit den meisten einmaligen Rektionskomposita sind in Tabelle 5 neben dem Anteil an syntaktisch belegten Fällen aufgeführt.

RK-Kopf	RK-Hapax Legomena	Als OV belegt	OV/RK
<i>Hersteller</i>	1130	92	0.081416
<i>Leiter</i>	1057	51	0.04825
<i>Führer</i>	867	147	0.16955
<i>Besitzer</i>	802	178	0.221945
<i>Anbieter</i>	716	136	0.189944
<i>Vertreter</i>	664	71	0.106928
<i>Macher</i>	629	240	0.381558
<i>Betreiber</i>	568	57	0.100352
<i>Lehrer</i>	392	30	0.076531
<i>Bewohner</i>	381	12	0.031496
<i>Sender</i>	366	21	0.057377
<i>Sammler</i>	344	1	0.002907

Tab. 5: Die meistbelegten Köpfe der als Hapax Legomena vorkommenden Rektionskomposita und ihr Verhältnis zu den OV-Paaren

Die Tabelle zeigt, dass der Anteil von syntaktisch belegten Kombinationen erstaunlich gering ist (in den meisten Fällen unter 10 %), obwohl das Muster OV sehr viel häufiger ist als die Bildung von Rektionskomposita. In einigen Fällen liegt dies an einer partiellen Suppletion mit anderen Verben, wie bspw. *Besitzer* neben *haben* (anstatt *Haben*) oder *Lehrer* neben *unterrichten* (anstatt *Unterrichten*). In anderen Fällen gibt es jedoch keinen prinzipiellen Grund für die Verteilung der Daten. Offenbar bildet man sehr viele Bezeichnungen für *Sammler*, deren Sammeltätigkeiten mit den entsprechenden Objekten im Korpus als verbale Phrasen nicht realisiert werden. Die Kopflexeme in der Tabelle sind natürlich stark lexikalisiert, aber die einzelnen Kombinationen hinter den Daten (bspw. *Suppenlöffelsammler* oder in einem übertragenen Sinn von „sammeln“ auch *Sonnenenergiesammler*) sind dennoch sicherlich produktive, nicht lexikalisierte Bildungen. Die Köpfe selbst tragen eine für -er-Nominalisierungen typische Berufsbezeichnungsbedeutung, die sich für die nominale Realisierung besonders gut eignet. Die große Zahl an produktiven Mustern mit diesen Köpfen, die syntaktisch nicht belegt sind, legt nahe, dass die einmal belegten Komposita nach einem lexikalischen Muster gebildet werden, das als Kopf das jeweilige deverbale Agensnomen spezifiziert.

Solche lexikalischen Muster entstehen durch einen Prozess von lexikalischer Verwurzelung („Entrenchment“), der Dank der hohen Frequenz

und Lexikalisierung des deverbalen Kopflexems zustande kommt. Das deverbale Nomen kann dann als eigenständiges relationales Nomen mit der dazugehörigen Argumentstruktur angesehen werden. So würden etwa neue Komposita auf *-leiter* entstehen, weil sich dieser Kopf als passend für die Beschreibung einer Person etabliert hat, die für die Leitung einer Organisation o.ä. zuständig ist, und nicht notwendigerweise über einen syntaktischen Zwischenschritt, der sich zumindest in den Fällen in Tabelle 4 nicht belegen lässt.

Gleichzeitig kann man auch spezielle Eigenschaften der RK-Konstruktion feststellen, die in der verbalen Variante nicht erscheinen und von dem strengen deverbalen Argumentmuster abweichen. Hier können wir nur kurz auf Prozesse von metonymischer Übertragung bzw. Ellipse eingehen (vgl. Gaeta & Zeldes Ms. für eine ausführlichere Diskussion), die zu einem semantischen Wandel führen. Im ersten Fall finden wir Beispiele, in denen das Rektionskompositum nicht das syntaktische direkte Objekt als Argument regiert, sondern einen indirekten Stellvertreter davon: Komposita wie *Besuchsleiter* oder *Erotikhersteller* finden eine plausible phrasale Entsprechung, wenn der Nicht-Kopf metonymisch verstanden wird, d.h. anstatt *?einen Besuch leiten* bzw. *?Erotik herstellen* würde die verbale Entsprechung eher *eine Besuchergruppe leiten* bzw. *Erotikartikel herstellen* lauten.¹⁸ Solche Prozesse sind generell charakteristisch für Komposition im Deutschen, wie ein vergleichbares Grounding-Kompositum wie *Erotikladen* zeigt, das als abgekürzte Form des ebenfalls belegten Wortes *Erotikartikelladen* verstanden werden kann.

4 Fazit

Die Verfügbarkeit eines sehr großen, wenn auch unausgewogenen, Korpus hat uns empirische Evidenz für die komplexen Verhältnisse zwischen Komposition und syntaktischen Fügungen im Deutschen geliefert. Die ungleichmäßigen Verteilungen der Lexeme und der Relationsarten zwischen Kompositionsgliedern zeigen uns, dass das mentale Lexikon eine bedeutende Rolle sowohl in der Sprachproduktion als auch in der Analyse von Komposita spielt. Wiederkehrende syntaktische Muster, die aufgrund ihrer Frequenz eine Spur im mentalen Lexikon hinterlassen, können die korrekte Interpretation der Wortbildungsbeutung in der Komposition

18 Wahrscheinlich lässt sich auch der Fall der Typen [N]-Vertreter wie *Staubsaugervertreter*, *Handelsvertreter* usw. mithilfe einer ähnlichen Ellipse erklären: *(eine) Staubsauger[firma] vertreten* > *Staubsaugervertreter*. Für eine eingehende Analyse s. Gaeta & Zeldes (Ms.).

vorhersagen. Dieser Befund widerspricht jedoch nicht der von vielen Forschern vertretenen Idee, dass kontextuell un(ter)spezifizierte semantische Informationen die Interpretation der Komposita lenken (vgl. u.a. Olsen in diesem Band). Wiederkehrende Ausdrücke spiegeln unsere pragmatische Erfahrung wider und bieten ein sprachliches Modell für die vorbestimmten Deutungen vieler Kompositareihen.

Auch für Rektionskomposita spielt die sprachliche Erfahrung eine zentrale Rolle, da ihre Erweiterung offensichtlich zumindest teilweise auf lexikalischen Mustern basiert, die nur indirekt eine syntaktische Struktur wiedergeben (hierfür muss man nicht unbedingt eine synchrone „*syntax below zero*“ annehmen, vgl. auch den „*Construction Morphology*“-Ansatz von Booij 2005, 2009). Nach diesen Mustern können neue Komposita analog gebildet werden, wie die Vielfältigkeit der syntaktisch nichtbelegten Hapax Legomena zeigt.

Das Verhältnis zwischen Syntax und Komposition wird durch unsere empirische Analyse nur teilweise bestätigt: Häufige syntaktische Muster können neue Komposita motivieren, es gibt aber auch häufige Komposita, die scheinbar kein syntaktisches Vorbild haben und in der Lage sind, neue Komposita zu motivieren oder ggf. auch einer sekundären syntaktischen Realisierung zugrunde liegen. Wir nehmen an, dass jede häufige Struktur eine Chance hat, eine selbstständige Darstellung im mentalen Lexikon zu erlangen, die dann Wortbildungsprozesse beeinflussen kann. Elektronische Korpora stellen dabei ein unentbehrliches Arbeitsmittel dar, das nicht nur authentische Daten zur qualitativen Untersuchung bereitstellt, sondern auch als erste quantitative Annäherung für den Inhalt des mentalen Lexikons verwendet werden kann.

5 Literatur

- Baayen, Harald (2001): *Word-Frequency Distributions*. Dordrecht: Kluwer.
 Baayen, R. Harald & Antoinette Renouf (1996): „Chronicle the Times: Productive Lexical Innovations in an English Newspaper“. *Language* 72/1, 69-96.
 Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta (2009): „The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora“. *Language Resources and Evaluation* 43/3, 209-26.
 Baroni, Marco, Emiliano Guevara & Vito Pirrelli (2007): „NN compounds in Italian: Modelling category induction and analogical extension“. In: Vito Pirrelli (Hg.), *Psycho-Computational Issues in Morphology Learning and Processing*. Sonderheft von *Lingue e Linguaggio* 6/2, 263-90.

- Baroni, Marco, Emiliano Guevara & Vito Pirrelli (2009): „Sulla tipologia dei composti N+N in italiano: Principi categoriali ed evidenza distribuzionale a confronto“. In: Ruben Benatti, Giacomo Ferrari & Monica Mosca (Hgg.), *Linguistica e modelli tecnologici di ricerca. Atti del 40esimo Congresso della Società di Linguistica Italiana*. Roma: Bulzoni, 73-95.
- Bauer, Laurie (2001): „Compounding“. In: Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (Hgg.), *Language Typology and Language Universals*. Berlin & New York: Mouton de Gruyter, 695-707.
- Booij, Geert E. (2005): „Compounding and Derivation: Evidence for Construction Morphology“. In: Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Franz Rainer (Hgg.), *Morphology and its Demarcations*. Amsterdam & Philadelphia: John Benjamins, 109-32.
- Booij, Geert E. (2009): „Compounding and Construction Morphology“. In: Rochelle Lieber & Pavol Štekauer (Hgg.), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 201-16.
- Downing, Pamela A. (1977): „On the Creation and Use of English Compound Nouns“. *Language* 53/4, 810-42.
- Firth, John R. (1957): *Papers in Linguistics*. London: Oxford University Press.
- Gaeta, Livio (2008): „Constituent order in compounds and syntax: typology and diachrony“. *Morphology* 18/2, 117-41.
- Gaeta, Livio (2010): „Synthetic compounds. With special reference to German“. In: Sergio Scalise & Irene Vogel (Hgg.), *Cross-Disciplinary Issues in Compounding*. Amsterdam & Philadelphia: John Benjamins, 219-36.
- Gaeta, Livio & Davide Ricca (2006): „Productivity in Italian word formation: A variable-corpus approach“. *Linguistics* 44/1, 57-89.
- Gaeta, Livio & Davide Ricca (2009): „*Composita solvantur*: Compounds as lexical units or morphological objects?“. In: Livio Gaeta & Maria Grossmann (Hgg.), *Compounds between syntax and lexicon*. Sonderheft von *Italian Journal of Linguistics* 21/1, 35-70.
- Gaeta, Livio & Zeldes, Amir (Ms.). *Are synthetic compounds syntactically motivated? The role of the lexicon on the border between syntax and morphology*, Manuskript, Humboldt Universität zu Berlin.
- Gallo, Francesca G. (2006-2007): *La composizione in tedesco: studi su un corpus del progetto WaCky*. Magisterarbeit. Università di Napoli "Federico II".
- Goldberg, Adele E. (2006): *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Harris, Zellig S. (1954): „Distributional Structure“. *Word* 10/2-3: 146-62.
- Heringer, Hans Jürgen (1984): „Wortbildung: Sinn aus dem Chaos“. *Deutsche Sprache* 12, 1-13.
- Lauer, Mark (1995): *Designing Statistical Language Learners: Experiments on noun compounds*. PhD. Diss. Macquarie University.
- Olsen, Susan (in diesem Band): „Der Einfluss des Mentalen Lexikons auf die Interpretation von Komposita“.
- Olsen, Susan (2001): „Copulative compounds. A closer look at the interface between Morphology and Syntax“. In: Geert Booij & Jaap van Marle (Hgg.), *Yearbook of Morphology 2000*. Dordrecht: Kluwer, 279-320.

- Pirrelli, Vito, Emiliano Guevara & Marco Baroni (2010): „Computational issues in compound parsing“. In: Sergio Scalise & Irene Vogel (Hgg.), *Cross-disciplinary issues in compounding*. Amsterdam & Philadelphia: Benjamins, 271-86.
- Roeper, Thomas & Siegel, Muffy E. A. (1978): „A Lexical Transformation for Verbal Compounds“. *Linguistic Inquiry* 9/2, 199-260.
- Siebert, Susann (1999): *Wortbildung und Grammatik. Syntaktische Restriktionen in der Struktur komplexer Wörter*. Tübingen: Niemeyer.
- Zeldes, Amir (2009): „Quantifying Constructional Productivity with Unseen Slot Members“. In: *Proceedings of the NAACL HLT Workshop on Computational Approaches to Linguistic Creativity, June 5, Boulder CO*. Stroudsburg, PA: Association for Computational Linguistics, 47-54.